STEMscopes NEW MEXICO 3D
K-12 SCIENCE
POWERED BY ACCELERATE LEARNING

A Comparison of 2023 NM-ASR 5th Grade Science Achievement for STEMscopes and Non-STEMscopes Schools in New Mexico

BACKGROUND

The efficacy of the STEMscopes Science 5th grade curriculum is examined in this study, by employing a post-facto quasi-experimental design (QED) with a matched control group, to test the potential association between STEMscopes Science and science outcomes on the 2023 New Mexico Assessment of Science Readiness (NM-ASR). **This approach, that is, QED efficacy studies with matching, meets the requirements for ESSA Tier 2 evidence as well as for the What Works Clearinghouse (WWC) 5.0 Group Design Standards with reservations. This is our first statewide evaluation in New Mexico.**

We also consider school-level science achievement in several student subpopulations. Specifically, previous studies (Morgan, Farkas, Hillemeier, & Maczuga, 2016) suggest there is a need for science curriculums that can engage students with diverse experiences and backgrounds and give them opportunities based on where they currently are in their science learning trajectory. Past STEMscopes reports suggest that STEMscopes may have a stronger association with standardized science test outcomes among minorities (particularly Black/African American students and Latinx/Hispanic students) as well as students considered low-income and students who receive special education services.

Thus, within the current report we consider two specific areas of enquiry. 1) We examine potential group differences in science achievement for STEMscopes versus non-STEMscopes schools. Overall, we hypothesize that even with the more stringent matched control group QED design, schools that purchased and used STEMscopes during the 2022–2023 school year will achieve higher science outcomes on the 2023 NM-ASR science test than schools that did not purchase STEMscopes (designated “non-STEMscopes schools”). Specifically, we tested whether a higher percentage of students “passed” (i.e., achieved scores of proficient or above) in STEMscopes versus non-STEMscopes schools. 2) We also anticipate significant associations between STEMscopes Science and school 5th grade NM-ASR science outcomes among subgroups of students.

RESULTS

To examine our hypothesis that STEMscopes Science schools will have higher 5th grade science outcomes on the NM-ASR, we conducted multiple regression analyses with 102 matched New Mexico schools. Our first set of analyses focused on predicting grade level rates of "passes"(includes students who scored Proficient or above) for 5th grade students taking the 2023 NM-ASR. We created a binary variable to indicate whether a school was a STEMscopes school or non-STEMscopes school, and we included several covariates (see Methods below). Results indicated a significant and positive increase in the school passing rate for the STEMscopes schools versus the non-STEMscopes schools (see Table 1 and Figure 1).

In addition, the subgroup analyses indicated significant, positive effects for all tested groups, including females, males, Hispanic/Latinx students, and students who were eligible for free and reduced lunch (please note that there was too much missing data to make conclusions about other student subgroups—see Methods below). Effects sizes ranged from 0.27 to 0.48, indicating medium to large effects.

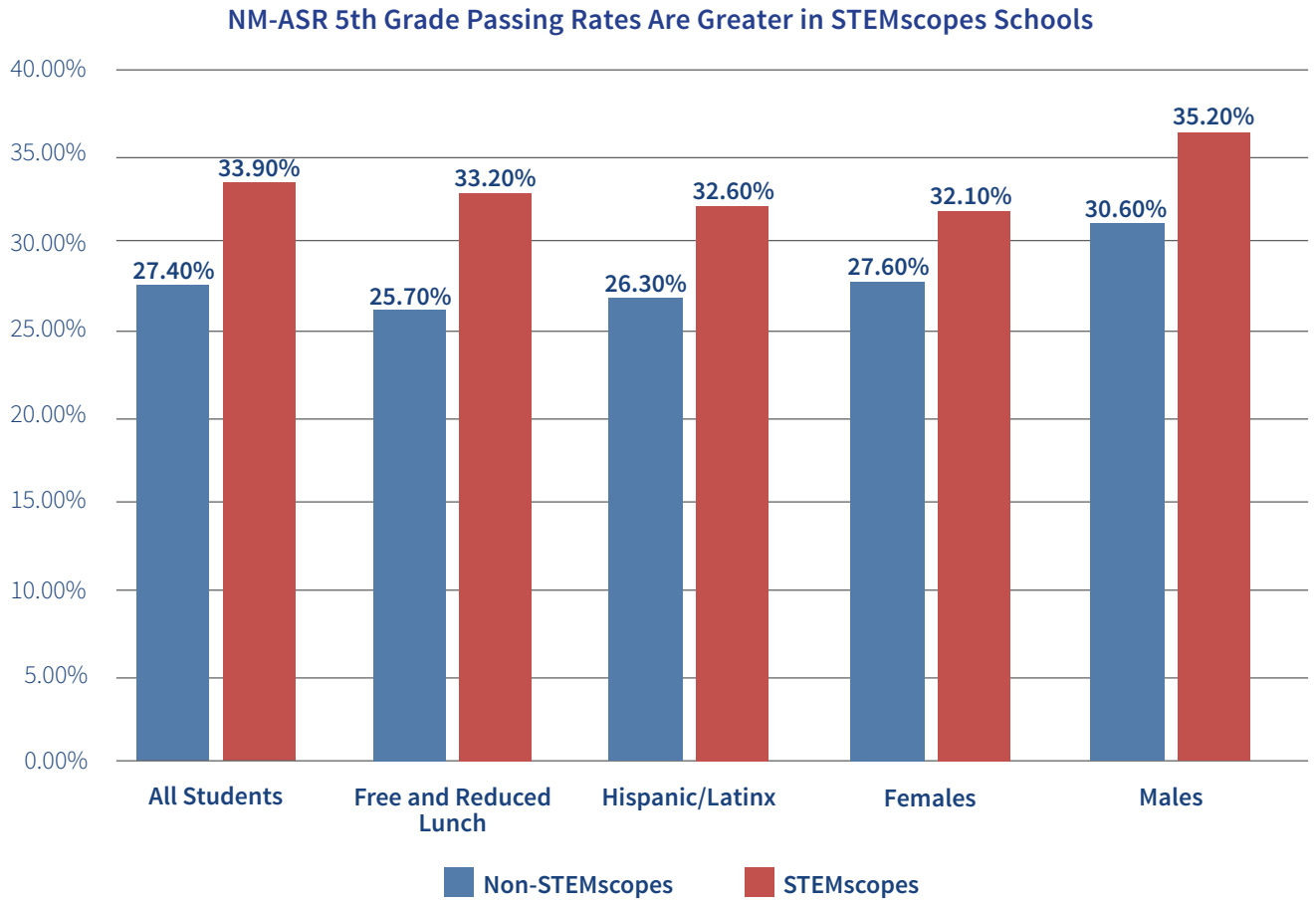

FIGURE 1
STEMscopes is positively related to NM-ASR school passing rates after controlling covariates.

METHODS

In this section we provide details about study procedures including the data sources, variables used, and participating schools.

Data sources

Data for this study came from two sources. First, we identified schools that purchased and used STEMscopes for 5th grade in the 2022–2023 school year through the STEMscopes analytics and Sales Force systems. Within the analytics reports, we used the number of 5th grade scopes accessed as a metric of use, and then confirmed usage with our internal Sales Force reports (to rule out schools that had free trial usage). Second, school demographic data and school performance on the NM-ASR were accessed through the New Mexico Public Education Department website. We used the 2021–2022 NM-ASR school achievement reports and employed the school level "Proficient and above" on the NM-ASR mathematics test as a baseline measure of academic achievement. Specifically, the state of New Mexico creates proficiency benchmarks in all academic content and identifies students as Novice, Nearing Proficiency, Proficient, and Advanced. The state of New Mexico uses the combined "Proficient and/or Advanced" as the state passing rate.

TABLE 1
Science Outcomes and Estimated Effects of STEMscopes Science (n=234)

Outcome	Non-STEMscopes M (SE)	STEMscopes M (SE)	STEMscopes vs Non-STEMscopes difference (b)	Effect Size (Hedge's G)
School level percent "Passed" NM-ASR (all students)	**27.40 (1.50)**	**33.90 (2.10)**	**6.50****	**0.39**
Male "Passed" NM-ASR	30.60 (1.80)	35.20 (2.30)	4.60*	0.28
Female "Passed" NM-ASR	27.60 (1.50)	32.10 (2.30)	4.50*	0.27
Low income*** "Passed" NM-ASR	25.70 (1.70)	33.20 (2.20)	7.50**	0.48
Hispanic "Passed" MCAS	26.30 (1.70)	32.60 (2.40)	6.30*	0.40

*= p < .05; **= p<.01; ***Low income as indexed by eligibility for free and reduced lunch.

We downloaded 2022–2023 school enrollment data by grade as well as subgroup data at the school level, which included enrollment percentages across race/ethnicity as well as other subpopulations, such as low-income students and students receiving special education services. We used these variables to match STEMscopes and non-STEMscopes schools (details are included in the Participants section below). Once matching was complete and baseline analyses were conducted (see Baseline Equivalence below), the 2023 NM-ASR school achievement reports for all students and for student sub-populations were downloaded. We analyzed school level data across the combined "Proficient and Advanced" benchmark used by the state.

Missing data

The New Mexico data associated with the Family Educational Rights and Privacy Act (FERPA) was characterized by some missing enrollment and subpopulation files. If less than 10% or more than 95% of students in a school belonged in a specific category, the state did not include the exact percentage but instead denoted the category <10 or >95. This led to missingness by design. The greatest impact of this lack of data appeared in some of the race/ethnicity categories. For example, 61% of information was missing from the variable "students of Asian descent," and it was unclear whether there were no students in this category enrolled in the school, or only a small number.However, most variables, such as baseline academic performance, were missing less than 5% of their data.

Given this limitation, we handled missing data in two ways. For any covariate variable used to match schools, we used multiple imputation by chained equations via R-studio's "MICE" package with the "CART" imputation method (see Van Buuren and Groothuis-Oudshoorn, 2011). We use MICE procedures during this step to ensure complete data for matching procedures via R-Studio's "Match-it" package. Once data were matched, in all final analyses we used R-Studio's "Lavaan" package, which uses full information maximum likelihood procedures to handle missing outcome data.

Participants

New Mexico is unique in the way they present public data: they do not separate school scores by grade. Thus if a school has grades K-8, then the NM-ASR will be an average passing rate across both the 5th and 8th grades. However, most New Mexico elementary schools only include preschool or kindergarten through 5th grade. As we wanted to isolate 5th grade achievement, we excluded schools that included both 5th and 8th grade levels. This resulted in 386 eligible schools. Of these schools, the overall number of elementary schools that purchased STEMscopes for the 2022–2023 school years was 63. Of these 63 schools, 51 used STEMscopes (in any capacity) for 5th grade (~14% of eligible schools).

To match schools based on the data available from the New Mexico Public Education Department, we used R-Studio's "Match-it" package, selecting one-to-one matching with "optimal" pair matching. We used 5th grade total enrollment (as an indicator of school size), school level race ethnicity percentages (including African Americans, White/Caucasian, Hispanic/Latinx, and Native American, as these categories had more complete data; the Asian and multi-racial subcategories were excluded due to greater than 50% missing), baseline achievement, and percentages of ELL

TABLE 2
Baseline Comparison of Matched STEMscopes and Non-STEMscopes Schools

Variables	State Total	Match Sample Total	Non-STEM-scopes	STEM-scopes	t-value	p-value	Effect Size
Baseline school 5th grade math passing rate 2022	27.7%	24.9%	24.1%	25.7%	0.61	0.55	0.12
Grade 5 enrollment	53.0	63.0	64.3	61.8	0.39	0.70	0.08
Percent low-income students	44.5%	51.3%	51.9%	50.7%	0.50	0.62	0.10
Percent Black/African American students	3.1%	3.1%	3.0%	3.1%	0.18	0.86	0.04
Percent Latinx/Hispanic students	65.2%	61.9%	59.4%	64.4%	0.94	0.35	0.19
Percent White/Caucasian students	23.7%	19.9%	19.6%	20.2%	0.16	0.87	0.03
Percent Native American students	17.5%	22.3%	25.3%	19.3%	0.97	0.33	0.19
Percent of English Language Learners (ELLs)	21.0%	18.0%	18.6%	17.4%	0.49	0.63	0.10
Percent of special education students	17.8%	16.9%	17.3%	16.5%	0.57	0.57	0.11

(English Language Learners), low-income, and students with a disability to match schools. We also included a variable to indicate the district's location (rural, urban, suburban, or town) as part of the matching procedure only.

Baseline equivalence

For all covariate variables (the variables used for matching), including baseline academic achievement, there were no significant differences between matched groups (see Table 2). However, the What Works Clearinghouse (WWC) standards require that baseline differences for a single variable with a standard mean difference greater than 0.05 must be controlled for statistically. Following the recommendation of Stuart (2010), we included all covariates in the final analyses as a complementary approach to matching, and a more stringent test of effects. Inclusion of these covariates satisfies the WWC standard, as several variables had standard mean differences greater than or equal to 0.05.

Planned analyses

Analyses were conducted with using R-studio's Lavaan structural equation modeling package because this package includes estimation with full information maximum likelihood (FIML) to handle missing data. These procedures ensure that in the final analyses the estimation is not biased. Our main variables of interest were the 2023 5th grade science outcomes on the NM-ASR. In addition to the main outcome analyses, we ran follow-up regressions to evaluate NM-ASR outcomes for male, female, low-income, and Hispanic student subpopulations. There was too much data missing for other subpopulation variables for them to be considered. Of note, we also included schools' "subclass" designation from the MatchIt program as a clustering variable; as this is consider good practice when using matched data, as it estimates cluster robust standard errors (Greifer, 2021).

DISCUSSION AND CONCLUSION

This study provides evidence of the efficacy of the STEMscopes Science 5th grade curriculum. Specifically, positive significant effects were found, indicating that a school's use of STEMscopes Science results in higher overall average passing rates on the NM-ASR relative to non-STEMscopes schools. Among subgroups, analyses indicated that school passing rates for the studied student subpopulations were also impacted by STEMscopes. The effect sizes indicated medium to large program effects (Kraft, 2020). Indeed, a large effect size of 0.48 for low-income students demonstrates great promise in helping these students pass the NM-ASR: on average, low-income students in STEMScopes schools passed at a rate that was 7.5 percentage points higher than their peers in non-STEMscopes schools; representing a nearly 30% increase in the number of low-income students passing. One of the limitations of the current study is that it focuses on change at the school level, which is less nuanced to program effects than considering individual student changes in science knowledge and skills. Nevertheless, this study provides a stringent test of effects using a quasi-experimental design that meets the WWC 5.0 standards with reservations (What Works Clearinghouse, 2022) and ESSA Tier 2 evidence.

WORKS CITED

Greifer, N. (2021). MatchIt: getting started. The Comprehensive R Archive Network. Retrieved June, 2024.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241-253.

Morgan, P.L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Science achievement gaps begin very early, persist, and are largely explained by modifiable factors. *Educational Researcher*, 45, 18-35.

Rangel, V. S., Bell, E. R., Monroy, C., & Whitaker, J. R. (2015). Year two results from an evaluation of STEMscopes Science curriculum. https://www.academia.edu/10164195/STEMscopes_Evaluation_Year_3

Sanetti, L. M. H. & Collier-Meek, M. C. (2019). *Supporting successful interventions in schools: Tools to plan, evaluate, and sustain effective implementation.* Guilford Press.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 25, 1.

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*, 1-67.

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook,* version 5.0. U.S. Department of Education, Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5_0-0-508.pdf